

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-231187
 (43)Date of publication of application : 05.09.1997

(51)Int.Cl.

G06F 15/173

(21)Application number : 08-034780
 (22)Date of filing : 22.02.1996

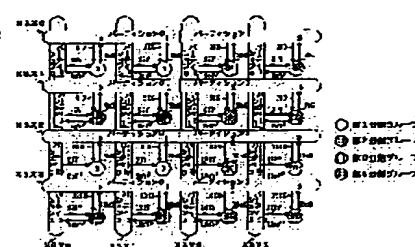
(71)Applicant : HITACHI LTD
 (72)Inventor : TSUJI KENJI
 ISHII MASAHIKO
 PATORITSUKU HAMIRUTON
 FUNAKUBO NOBUO

(54) DATA TRANSFER METHOD OF PARALLEL COMPUTER

(57)Abstract:

PROBLEM TO BE SOLVED: To guarantee no transfer of packets exceeding a divided group at the time of divisional operation of the parallel computer.

SOLUTION: Crossbar switches XBX0-XBX3, and XBY0-XBY3 are divided into two partitions 0 and 1. In this case, nodes are divided into a 1st group of 0, 1, 8, and 9, a 2nd group of 2, 3, 10, and 11, a 3rd group of 4, 5, 12, and 13, and a 4th group of 6, 7, 14, and 15. Each node holds port information which belongs to the corresponding partition of the XB to which the node is connected and each XB holds port information which belongs to its partitions. Each node and each XB compare the previously held port information with the routing information of a packet when transferring the packet, and send the packet out only when they match each other.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C) 1998,2000 Japanese Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-231187

(43)公開日 平成9年(1997)9月5日

(51)Int.Cl.⁶

G 0 6 F 15/173

識別記号

庁内整理番号

F I

G 0 6 F 15/16

技術表示箇所

4 0 0 W

審査請求 未請求 請求項の数3 OL (全10頁)

(21)出願番号 特願平8-34780

(22)出願日 平成8年(1996)2月22日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 辻 勝司

神奈川県秦野市堀山下1番地 株式会社日立製作所汎用コンピュータ事業部内

(72)発明者 石井 将人

神奈川県秦野市堀山下1番地 株式会社日立製作所汎用コンピュータ事業部内

(72)発明者 パトリック ハミルトン

神奈川県秦野市堀山下1番地 株式会社日立製作所汎用コンピュータ事業部内

(74)代理人 弁理士 鈴木 誠

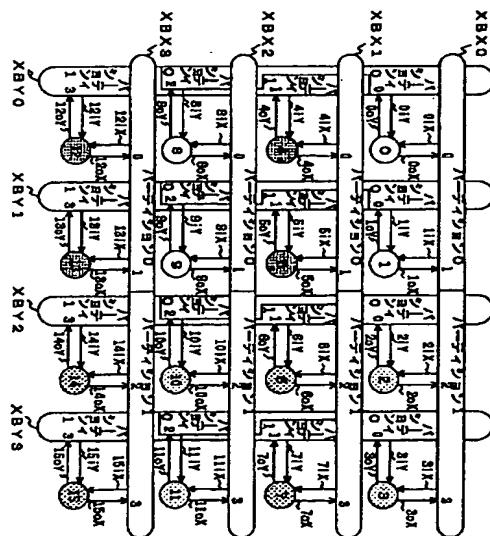
最終頁に続く

(54)【発明の名称】並列計算機のデータ転送方法

(57)【要約】

【課題】並列計算機の分割運転時、パケットが分割グループを越えて転送されないことを保証する。

【解決手段】クロスバスイッチXBX0～XBX3, XBY0～XBY3を各々パーティション0, 1に2分割する。この場合、ノードは、0, 1, 8, 9の第1グループ、2, 3, 10, 11の第2グループ、4, 5, 12, 13の第3グループ、6, 7, 14, 15の第4グループに分かれる。各ノードには自ノードが接続されるXBの該当パーティションに属するポート情報を、各XBには自XBの各パーティションに属するポート情報を各々保持せしめる。各ノード、各XBは、パケットを転送する際、該パケットのルーティング情報をとらかじめ保持されたポート情報を比較し、照合がとれた場合のみ該パケットを送出する。



○ 第1分割グループ
■ 第2分割グループ
△ 第3分割グループ
◆ 第4分割グループ

【特許請求の範囲】

【請求項1】 複数個のプロセッサ（以下、ノード）というと、複数個の入出力ポートを持ったクロスバスイッチからなる相互結合網で構成された並列計算機において、ノードを複数のグループに分割し、同一グループに属するノード同士でパケット形式によりデータを転送する方法であって、

クロスバスイッチの入出力ポートを複数の部分に分割し（以下、分割した各部分をパーティションという）、各ノードに、自ノードが接続されるクロスバスイッチの該当パーティションに属する出力ポート情報をあらかじめ保持し、

各ノードは、パケットを転送する前に、該パケットにルーティング情報として付加されている転送先ノードのポート情報を前記あらかじめ保持された出力ポート情報を照合し、照合がとれたならば、自ノードが接続されるクロスバスイッチに該パケットを送出し、照合がとれない場合には該パケットの送出を中止することを特徴とする並列計算機のデータ転送方法。

【請求項2】 複数個のプロセッサ（以下、ノードという）と、複数個の入出力ポートを持ったクロスバスイッチからなる相互結合網で構成された並列計算機において、ノードを複数のグループに分割し、同一グループに属するノード同士でパケット形式によりデータを転送する方法であって、

クロスバスイッチの入出力ポートを複数の部分に分割し（以下、分割した各部分をパーティションという）、各クロスバスイッチに、自クロスバスイッチの各パーティションに属する入出力ポート情報をあらかじめ保持し、

各クロスバスイッチは、パケットを入力ポートで受信すると、入力ポート情報により該入力ポートの属するパーティションを選択し、前記受信したパケットにルーティング情報として付加されている転送先ノードのポート情報を前記選択されたパーティションに属する出力ポート情報と照合し、照合がとれたならば、該当出力ポートより該パケットを送出し、照合がとれない場合には該パケットの送出を中止することを特徴とする並列計算機のデータ転送方法。

【請求項3】 請求項1、2記載の並列計算機のデータ転送方法において、分割グループの構成変更に応じ、各ノードの出力ポート情報あるいは各クロスバスイッチの入出力ポート情報の内容を変更することを特徴とする並列計算機のデータ転送方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、並列計算機のデータ転送方法に係り、特に並列計算機の分割運転において、分割を行ったノードグループを越えてパケットが転送されないことを保証し、分割したノードグループが相

互に影響を与えない高信頼性を要求される並列計算機のデータ転送方法に関する。

【0002】

【従来の技術】 従来、複数プロセッサのノード群と該ノード群をクロスバスイッチにより相互結合した構成の並列計算機において、種々の利用目的等に応じてノード群を複数グループに分割し、各グループ毎に分割運転を行う場合、一般にパケットに宛先ノード情報の他にグループ識別情報を付加することで、同一グループに属するノード同士のデータ転送を実現していた。なお、並列計算機の分割運転に関する公知例としては、例えば特開平5-28122号公報が挙げられる。

【0003】

【発明が解決しようとする課題】 従来技術においては、転送元ノードがプログラム（OS）のバグ等により転送先ノードを誤り、転送パケットに他の分割グループ内のノード等を指定した場合、あるいは、転送途中において信号ディレーヤノイズによるビット欠け・ビット化けにより、パケットの宛先ノードやグループ情報が変化した場合、誤って他の分割グループにパケットが転送されるという問題があった。

【0004】 本発明の目的は、並列計算機の分割運転時、誤って他の分割グループにパケットが転送されないことを保証し、分割グループ間が相互に影響することなく運用できる、高信頼性のデータ転送方法を提供することにある。

【0005】

【課題を解決するための手段】 本発明は、複数個のプロセッサ（ノード）と、複数個の入出力ポートを持ったクロスバスイッチからなる相互結合網で構成された並列計算機において、ノードを複数のグループに分割し、同一グループに属するノード同士でパケット形式によりデータを転送する際、クロスバスイッチの入出力ポートを複数の部分に分割し（分割した各部分をパーティションという）、各ノードに、自ノードが接続されるクロスバスイッチの該当パーティションに属する出力ポート情報をあらかじめ保持しておく。各ノードは、パケットを転送する前に、該パケットにルーティング情報として付加されている転送先ノードのポート情報を前記あらかじめ保持された出力ポート情報と照合し、照合がとれたならば、自ノードが接続されるクロスバスイッチに該パケットを送出し、照合がとれない場合には該パケットの送出を中止する。

【0006】 また、本発明は、各クロスバスイッチに、自クロスバスイッチの各パーティションに属する入出力ポート情報をあらかじめ保持し、各クロスバスイッチは、パケットを入力ポートで受信すると、まず、入力ポート情報により該入力ポートの属するパーティションを選択する。次に、受信したパケットにルーティング情報として付加されている転送先ノードのポート情報を前記

選択されたパーティションに属する出力ポート情報と照合し、照合がとれたならば、該当出力ポートより該パケットを送出し、照合がとれない場合には該パケットの送出を中止する。

【0007】これにより、例えば信号ディレー、ノイズによるビット欠け・ビット化や、OSのバグ等により転送先ノードを誤り他の分割グループ内のノードを指定した場合等に、エラーを検出でき、他の分割グループに影響を与える分割運転時の信頼性を向上できる。

【0008】

【発明の実施の形態】以下、本発明の一実施例について、16個のノード（プロセッサ）が、X方向に4個、Y方向に4個のクロスバスイッチからなる相互結合網で接続される並列計算機を用いて説明する。

【0009】図1は、本実施例の16個のノードと、それらのノードを相互結合するX方向に4個、Y方向に4個のクロスバスイッチ（XB）からなる並列計算機を示す。図1において、0～15はノード（プロセッサ）、 $XB_X 0 \sim XB_X 3$ はX方向のクロスバスイッチ（XB_X）、 $XB_Y 0 \sim XB_Y 3$ はY方向のクロスバスイッチ（XB_Y）である。本実施例に用いる各クロスバスイッチは、入出力ポートを各々4個持つておらず、ここではそれらに0～3のポート番号を付ける。ノードとクロスバスイッチ間は、各々、入力信号線 $0iX \sim 15iX$ 、 $0iY \sim 15iY$ 、出力信号線 $0oX \sim 15oX$ 、 $0oY \sim 15oY$ で接続される。ただし、この入出力信号線の標記は、ノード側からみた場合であり、クロスバスイッチ側からは入出力の関係が逆になる。

【0010】図2は、本実施例の相互結合網におけるノード間通信の通信パケットを表す。200はパケット全体を表わし、210と220はルーティングフィールド、250はデータフィールドである。ルーティングフィールド210、220はパケットの先頭に置かれ、転送先ノードのポート情報として、210には各XB_Xの選択すべき出力ポートを表す情報が格納され、220には各XB_Xの選択すべき出力ポートを示す情報が格納される。250には転送すべきデータが格納される。

【0011】本実施例では、XB_Xルーティングフィールド210には、転送先ノード番号を4ビットの2進数で表わした場合の下位2ビットをデコードした値（4ビット）即ち、XB_Xの出力ポート番号を示す値が、また、XB_Yルーティングフィールド220には、同様に転送先ノード番号を2進数で表わした場合の4ビット中の上位2ビットをデコードした値（4ビット）、即ち、XB_Yの出力ポート番号を示す値が格納される。図3に、ノード0～15に対応するルーティングフィールド210、220の値を示す。

【0012】例えば、ノード1からノード7へパケットを転送する際、ノード1では、XB_Xルーティングフィールド210にXB_X出力ポート番号3を示す“000

1”を、また、XB_Yルーティングフィールド220にXB_Y出力ポート番号1を示す“0100”を各々格納して、例えば、出力信号線 $1oX$ よりXB_X0へ送出する。なお、XB_X0へ送出するか、XB_Y1へ送出するかは、乱数や結合網の通信負荷バランス、もしくはそれらの組み合わせで決定される。

【0013】ここで、分割運転を実施しない場合、XB_X0では、ノード1が送出したパケットを入力ポート1より受信し、該パケットのXB_Xルーティングフィールド210の“0001”に従って、出力ポート3より該パケットを送出する。該パケットをノード3が入力信号線 $3iX$ より受信し、XB_X／XB_Yにルーティングフィールドより自分宛でないことを確認すると、該パケットを出力信号線 $3oY$ よりXB_Y3へ送出する。XB_Y3では、該パケットを入力ポート0より受信し、該パケットのXB_Yルーティングフィールド220の“0100”に従って、出力ポート1より該パケットを送出する。該パケットをノード7が入力線 $7iY$ より受信し、XB_X／XB_Yルーティングフィールドより自分宛であることを確認し、該パケットを取り込む。

【0014】これが分割運転を実施しない場合の一般的動作であるが、以下に、図1の実施例の分割運転について説明する。

【0015】図1において、ノード0、1、8、9を第1分割グループ、ノード2、3、10、11を第2分割グループ、ノード4、5、12、13を第3分割グループ、ノード6、7、14、15を第4分割グループとして、計4グループに分割し、各グループ単位で分割運転を行うとする。このため、クロスバスイッチを各々論理的にパーティション0、パーティション1に2分割する。この場合、図1に示すように、X方向クロスバスイッチXB_X0～XB_X3では、ポート0と1をパーティション0、ポート2と3をパーティション1に割り当て、Y方向クロスバスイッチXB_Y0～XB_Y3では、ポート0と2をパーティション0、ポート1と3をパーティション1に割り当てる。図4に、各ノードの分割グループとXB_X、XB_Yのパーティションの対応関係を示す。

【0016】後述するように、各ノードは、自ノードが属する分割グループに対応する、XB_X、XB_Yの各パーティションに属する出力ポート情報（出力可能ポート情報）を分割情報として格納する分割情報レジスタを具備する。また、各クロスバスイッチは、自クロスバスイッチを論理的に分割する各パーティションに属する入出力ポート情報を分割構成情報として格納する構成情報レジスタを具備する。分割運転を行う前に、各グループのノードの分割情報レジスタ、各クロスバスイッチの構成情報レジスタに当該情報を格納しておく。

【0017】分割運転時、各ノード、各クロスバスイッチでは、パケットを送出するに先立ち、該パケットのル

ーティングフィールドの内容を該当レジスタにあらかじめ格納されている情報と照合して、パケット送出の可否を確認する。その結果、送出可であれば、パケットを送出し、不可であれば、パケットの送出を中止する。これにより、各分割グループ、即ち、 XB_X 、 XB_Y の各パーティションのグループに属するノード群ごとのデータ転送が実現する。

【0018】以下、各ノード、各クロスバスイッチの具体的構成例について説明する。図5は、本実施例のノードの構成例を示す。ここでは、図1中のノード1を例にとるが、他のノードも同様の構成である。図5において、100は演算ユニット、105はローカルメモリ、115は通信ユニットであり、これらは内部バス110で相互接続される。該ノード1には、入力信号線 i_X 、 i_Y に各々接続される入力ポート i_X 、 i_Y 、及び、出力信号線 o_X 、 o_Y に各々接続される出力ポート o_X 、 o_Y がある。入力ポート i_X 、 i_Y は各々信号線111、112で通信ユニット115に接続され、通信ユニット115は各々信号線113、114で出力ポート o_X 、 o_Y に接続される。120は分割情報レジスタ、121はエラー情報レジスタ121であり、通信ユニット115に設けられる。

【0019】演算ユニット100は、あるノードにデータを転送する際、先頭にルーティングフィールドを付加したパケットをローカルメモリ105に生成する。通信ユニット115は、該ローカルメモリ105のパケットを出力ポート o_X あるいは o_Y から送出する。また、通信ユニット115は、入力ポート i_X あるいは i_Y で受信したパケットが自ノード宛の場合、該パケットをローカルメモリ105へ転送する。一方、自ノード宛でない場合、入力ポート i_X で受信したパケットは出力ポート o_Y へ送出し、入力ポート i_Y で受信したパケットは出力ポート o_X へ送出する。

【0020】ここで、入力ポート i_X 、 i_Y で受信したパケットあるいは自ノードで生成したパケットを出力ポート o_X 、 o_Y に送出する際、通信ユニット115にて、該パケットのルーティングフィールドの内容を分割情報レジスタ120の分割情報と照合して、パケット送出の可否を確認する。その結果、不可であれば、該パケットの送出を中止し、エラー情報レジスタ121をオンとする。

【0021】図6は、分割情報を格納する分割情報レジスタ120の構成例を示す。該レジスタ120に格納される分割情報は、当該ノードが接続される XB_X 、 XB_Y の各パーティションの出力ポート情報である。この XB_X 、 XB_Y の当該パーティションに属する出力ポートは、該 XB_X 、 XB_Y に接続されるノードから入力したパケットの送出が許可されるポート（出力可能ポート）を意味している。

【0022】本実施例では、分割情報レジスタ120は

8ビットからなる。該分割情報レジスタ120の上位4ビットは上位から順番に、 X 方向クロスバスイッチ（ XB_X ）の出力ポート0から3に対応しており、下位4ビットは上位から順番に、 Y 方向クロスバスイッチ（ XB_Y ）の出力ポート0から3に対応している。各ビットに1が設定されている場合は、そのビットに対応するクロスバスイッチの出力ポートからパケットを送出できることを示す。該分割情報レジスタ120に格納されている分割情報が等しいノード同士が同一の分割グループに属することになる。

【0023】図7に、図1に示す分割運転を行う場合の、各分割グループに属するノードの分割情報レジスタ120に格納される分割情報の具体的ビットパターンを示す。例えば、図1より、第1分割グループに属する各ノードは、各々、 $XB_X 0$ 、 $XB_X 2$ のパーティション0、 $XB_Y 0$ 、 $XB_Y 1$ のパーティション0のポートに接続されている。ここで、各 XB_X のパーティション0のポート番号は0と1、また、各 XB_Y のパーティション0のポート番号は0と2である。したがって、第1分割グループに属する各ノードの分割情報レジスタ120の、上位4ビットは“1100”、下位4ビットは“1010”であり、全体では“11001010”となる。同様にして、第2乃至第4分割グループに属する各ノードの分割情報レジスタ120の内容は、各々、“00111010”、“11000101”、“00110101”となる。

【0024】なお、分割運転を実施しない場合は、各ノードの分割情報レジスタ120の全ビットを“1”とすればよい。即ち、この場合は、 XB_X 、 XB_Y の全ての出力ポートからパケットの送出が許可されることを示している。

【0025】図8は、本実施例でノードを相互結合するクロスバスイッチの構成例を示す。ここでは、図1中の $XB_X 0$ を例にとるが、他のクロスバスイッチも同様の構成である。各クロスバスイッチは、入出力ポートを各々4個持つおり、順番に0～3のポート番号が付けられている。 $XB_X 0$ では、入力ポート0～3は各々入力信号線 $0i_0$ ～ $0i_3$ に接続され、出力ポート0～3は各々出力信号線 $0o_0$ ～ $0o_3$ に接続される。ここで、入力信号線 $0i_0$ ～ $0i_3$ 、出力信号線 $0o_0$ ～ $0o_3$ は、各々図1のノード0～3の出力信号線 $0o_X$ ～ $3o_X$ 、入力信号線 $0i_X$ ～ $3i_X$ に対応する。 130 は構成情報レジスタ、 131 はエラー情報レジスタである。

【0026】クロスバスイッチは、入力ポート0～3のいずれかに到来したパケットを、該パケットのルーティングフィールドの転送先ノード情報にもとづいて、出力ポート0～3のいずれかを選択して送出する。この時、該パケットのルーティングフィールドの内容を構成情報レジスタ130の分割構成情報と照合して、パケット送出の可否を確認する。その結果、不可であれば、該パケッ

トの送出を中止し、エラー情報レジスタ 131 をオンとする。

【0027】図9は、クロスバスイッチ内の分割構成情報を格納する構成情報レジスタ 130 の構成例を示す。該レジスタ 130 に格納される分割構成情報は、該クロスバスイッチを論理的に分割する各パーティションに属する入出力ポート情報である。パケット転送は、該レジスタ 130 の分割構成情報で示される同一パーティションに属する入力ポートと出力ポート間でのみ許可される。

【0028】本実施例では、図1に示す分割運転を行うため、各クロスバスイッチをパーティション0、パーティション1に2分割している。これに対応して、構成情報レジスタ 130 は、パーティション0に属する入力ポート構成情報を格納する入力ポート構成情報レジスタ (0) 131、出力ポート構成情報を格納する出力ポート構成情報レジスタ (0) 132、パーティション1に属する入力ポート情報構成を格納する入力ポート構成情報レジスタ (1) 133、出力ポート構成情報を格納する出力ポート構成情報レジスタ (1) 134 の計4個のレジスタからなる。各レジスタ 131～134 は4ビットからなる。これらレジスタのビットは上位から順番に、各々クロスバスイッチの入力ポート 0～3、出力ポート 0～3 に対応しており、各パーティションに属する入力ポートに対応するビットに“1”、出力ポートに対応するビットに“1”を各々格納し、各パーティション毎に、“1”的立っている入出力ポート間でパケットを転送できることを示す。

【0029】図10に、図1に示す分割運転を行う場合の、XB X 0～XB X 3、XB Y 0～XB Y 3 におけるパーティション0の各入／出力ポート構成情報をレジスタ 131、132、及び、パーティション1の各入／出力ポート構成情報をレジスタ 133、134 に格納される入／出力構成情報の具体的ビットパターンを示す。例えば、図1より、XB X 0 のパーティション0に属する入出力ポートはポート番号0と1である。したがって、入力ポート構成情報をレジスタ (0) 131 と出力ポート構成情報をレジスタ (0) 132 の内容は各々“1100”となり、XB X 0 のパーティション0では、入力ポート0、1のパケットを出力ポート0、1に送出できることを示している。また、XB X 0 のパーティション1に属する入出力ポートはポート番号2と3である。したがって、入力ポート構成情報をレジスタ (1) 133 と出力構成情報をレジスタ (1) 134 の内容は各々“0011”となり、XB X 0 のパーティション1では、入出力ポート2、3のパケットを出力ポート2、3に送出できることを示している。他のXB X、XB Y の各レジスタ 131～134 についても同様である。

【0030】なお、分割運転を実施しない場合は、各クロスバスイッチの構成情報をレジスタ 130 中のパーティ

ション0の入出力ポート構成情報をレジスタ 131、132 に全ビット“1”を格納し、パーティション1の入出力ポート構成情報をレジスタ 133、134 に全ビット“0”を格納する。即ち、各クロスバスイッチの論理分割を見掛上、パーティション0のみとするのである。これで、各クロスバスイッチが分割されず、全ての入力ポートからパケットを受信し、全ての出力ポートからパケットを送信できることになる。

【0031】次に、本実施例の分割運転時の動作について、同一の分割グループ内の任意のノード間の1対1通信を例に説明する。ここでは、図1において、第1分割グループ内のノード1からノード8へパケットを転送すると仮定する。分割運転を行う前に、あらかじめ各グループのノードの分割情報をレジスタ 120 には、図7に示す情報を、また、各クロスバスイッチの構成情報をレジスタ 130 の各レジスタ 131～134 には、図10に示す情報を格納しておく。

【0032】分割運転時、第1分割グループ内に属するノード1からノード8へパケットを転送する際、ノード1の演算ユニット 100 は、図2に示すパケット 200 として、図3より、XB X ルーティングフィールド 210 には、転送先ノード番号8を2進数で表わした“1000”的下位2ビットをデコードした値、即ち、XB X の出力ポート番号0を示す値“1000”を、XB Y ルーティングフィールド 220 には、同転送先ノード番号8を2進数で表わした“1000”的上位2ビットをデコードした値、即ち、XB Y の出力ポート番号2を示す値“0010”を、また、データフィールド 250 には転送データを各々格納し、該パケット 200 を内部バス 110 を通してローカルメモリ 105 に生成する。次に、演算ユニット 100 は、例えば乱数を発生させた結果、使用する出力ポートを0Xとし、内部バス 110 を通して通信ユニット 115 に、該パケット 200 の転送指示をする。通信ユニット 115 は、演算ユニット 100 からパケット 200 の転送指示を受けると、内部バス 110 を通して、ローカルメモリ 105 から転送すべきパケット 200 を読み出し、該パケット 200 のXB X ルーティングフィールド 210 の値“1000”から分割情報をレジスタ 120 の上位4ビット中の、XB X 出力ポート0と対応したビット(1ビット目)を参照し、該当ビットに“1”が格納されており、且つ、該パケット 200 のXB Y ルーティングフィールド 220 の値“0010”から分割情報をレジスタ 120 の下位4ビット中の、XB Y 出力ポート2と対応したビット(3ビット目)を参照し、該当ビットに“1”が格納されているかを確認する。本例では、図7より、ノード1の通信ユニット 115 内の分割情報をレジスタ 120 は“11001010”であり、その上位4ビット中のXB X 出力ポート0と対応したビット(1ビット目)は“1”、且つ、下位4ビット中のXB Y 出力ポート2と対応したビット

(3ビット目)も“1”であり、照合がとれる。そこで、通信ユニット115は、指定された出力ポート₀Xから出力信号線₁₀Xを通してXBX0に、該読み出したパケット200を転送する。

【0033】なお、ノード1の通信ユニット115は、転送パケットのルーティングフィールド210, 220の値から分割情報レジスタ120の該当出力ポートと対応した2ビットを参照して、何れか一つのビットにでも“1”が格納されていなかったならば、パケットの転送処理を中止し、内部バス110を介して演算ユニット100に割り込みを発生する。または、図5に示すエラー情報レジスタ121をオンとする。

【0034】XBX0は、ノード1からのパケット200を、入力ポート1が受け取る。該XBX0は、まず、図9に示す入力ポート構成情報レジスタ131, 133中の入力ポート1と対応したビットを参照し、該当ビットに“1”が格納されている側のパーティションを選択する。本例では、図10中のXBX0のレジスタ131, 133の内容より、パーティション0を選択する。次に、パケット200のXBXルーティングフィールド210の値“1000”を参照し、出力ポート0を選択する。次に、パーティション0の出力ポート構成情報レジスタ132の出力ポート0に対応するビットを参照し、該当ビットに“1”が格納されているか確認する。本例では、図10中のXBX0のレジスタ132の内容より、該当ビットに“1”が格納されているので、XBX0は、出力ポート0より出力信号線000を通して、該信号線000(入力信号線01X)と接続されているノード0にパケット200を転送する。

【0035】パケット200を受け取ったノード0の通信ユニットは、該パケットの転送先ノード番号を示すXBX/XBYルーティングフィールドが自ノード0宛でないことを確認し、さらに、ノード1の場合に、該ルーティングフィールドの値と該通信ユニット内の分割情報レジスタ120の値を照合して、照合がとれたことを確認した後、出力ポートYから信号線00Yを通してパケット200をXBY0に転送する。

【0036】XBY0は、ノード0からのパケット200を、入力ポート0が受け取ると、まず、該XBY0の入力ポート構成情報レジスタ131, 133の入力ポート0と対応したビット(1ビット目)を参照し、本例では、図10より、該当ビットに“1”が格納されているパーティション0を選択する。次に、パケット200のXBYルーティングフィールド200の値“0.010”を参照し、出力ポート2を選択する。次に、パーティション0の出力ポート構成情報レジスタ132の出力ポート2に対応するビット(3ビット目)を参照し、本例では、図10より、該当ビットに“1”が格納されているのを確認した後、出力ポート2から信号線802を通して、該信号線802(入力信号線81Y)と接続されてい

るノード8にパケット200を転送する。

【0037】パケット200を受け取ったノード8の通信ユニットは、該パケットの転送先ノード番号を示すルーティングフィールドが自ノード宛であることを確認すると、該パケットをローカルメモリへ取り込み、演算ユニットにパケット受信を通知する。

【0038】なお、クロスバススイッチXBX0, XBY0は、その構成情報レジスタの各入力/出力ポート構成情報レジスタのビットを参照した際に、何れか一つのビットにでも“1”が格納されていなければ、パケットの転送処理は中止し、エラー情報レジスタ131をオンとする。

【0039】次に、本実施例の分割運転において、同一の第1分割グループに属するノード1からノード8へパケットを転送する際に、OS等のエラーが発生し、ノード1が他の第3分割グループに属するノード4を転送先に指定した場合の動作を説明する。

【0040】ノード1の演算ユニット100は、パケットの転送先ノード番号8をノード番号4と誤って指定した場合、図2のパケット200として、図3より、XBXルーティングフィールド210には、ノード番号4を2進数で表わした“0100”的下位2ビットをデコードした値、即ち、XBXの出力ポート番号0を示す値“1000”を、XBYルーティングフィールド220には、同ノード番号“0100”的上位2ビットをデコードした値、即ち、XBYの出力ポート番号1を示す値“0100”を、データフィールド250には転送データを各々格納し、該パケット200を内部バス110を通してローカルメモリ105に生成する。次に演算ユニット100は、乱数を発生させた結果、例えば使用する出力ポートを₀Xとし、内部バス110を通して通信ユニット115に該パケット200の転送指示をする。通信ユニット115は、演算ユニット100からパケット200の転送指示を受けると、内部バス110を通して、ローカルメモリ105から転送すべきパケット200を読み出し、該パケット200のXBXルーティングフィールド210の値“1000”から分割情報レジスタ120の上位4ビット“1100”中の、XBX出力ポート0と対応したビット(1ビット目)を参照し、当該ビットに“1”が格納されていることを確認する。次に、パケット200のXBYルーティングフィールド220の値“0100”から分割情報レジスタ120の下位4ビット“1010”中の、XBY出力ポート1と対応したビット(2ビット目)を参照するが、当該ビットは“0”である。このため、ノード1の通信ユニット115は、パケットの転送処理を中止し、バス110を介して演算ユニット100に割り込みを発生させる。または、エラー情報レジスタ121をオンとする。

【0041】これにより、分割運転時、転送元ノードが誤って自分の属する分割グループとは別の分割グループ

のノードを転送先に指定した場合、該転送元ノードにて直ちにパケットの転送を抑止できる。

【0042】次に、本実施例の分割運転において、同一の第1分割グループ内に属するノード1からノード8へパケットを転送する際に、ノード1とXBX0間でノイズが発生し、ビット化けにより転送先ノード番号が8と12の両方に変わった場合の動作を説明する。

【0043】ノード1の演算ユニット100は、先と同様に、XBXルーティングフィールドに、転送先ノード番号8を2進数で表わした“1000”的下位2ビットをデコードした値、即ち、XBXの出力ポート番号0を示す値“1000”を、XBYルーティングフィールドに、同転送先ノード番号8を2進数で表わした“1000”的上位2ビットをデコードした値、即ち、XBYの出力ポート2を示す値“0010”を、転送データフィールドに転送データを各々格納し、図2に示されるパケット200を内部バス110を通してローカルメモリ105に生成する。次に、演算ユニット100は、例えば使用する出力ポートを0Xとし、内部バス110を通して通信ユニット115に該パケット200の転送指示をする。通信ユニット115は、演算ユニット100からパケット200の転送指示を受けると、内部バス110を通して、ローカルメモリ105から該パケット200を読み出し、パケット200のXBXルーティングフィールド210の値“1000”的上位4ビット中の、XBX出力ポート0と対応したビット(第1ビット目)を参照し、該当ビットに

“1”が格納されており、且つ、XBYルーティングフィールド220の値“0010”から分割情報レジスタ120の下位4ビット中の、XBY出力ポート2と対応したビットを参照し、該当ビットに“1”が格納されているのを確認した後、指定された出力ポート0Xから出力信号線10Xを通してXBX0に該パケット200を転送する。

【0044】ノード1からXBX0へパケット200を転送中にノイズが発生し、該パケット200のXBYルーティングフィールド220の値が“0010”から“0011”、即ち、XBYの出力ポートが2と3を示す値に代わったとする。

【0045】XBX0は、ノード1からのパケット200を、入力ポート1が受け取る。該XBX0は、入力ポート構成情報レジスタ131、133の入力ポート1と対応したビット(2ビット目)を参照し、図10より、該当ビットに1が格納されているパーティション0を選択する。次に、パケット200のXBXルーティングフィールド210の値“1000”を参照し、出力ポート0を選択する。次に、パーティション0の出力ポート構成情報レジスタ132の出力ポート0に対応するビット(1ビット目)を参照し、図10より、該当ビットに“1”が格納されているのを確認する。そこで、XBX

0は、出力ポート0より出力信号線000を通して、該信号線000(入力信号線0iX)と接続されているノード0にパケット200を転送する。

【0046】パケット200を受け取ったノード0の通信ユニットは、該パケット200の転送先ノード番号を示すXBX/XBYルーティングフィールドが自ノード0でないことを確認し、また、該ルーティングフィールドの値と自分割情報レジスタ120の値の照合がとれていることを確認した後、出力ポートYから出力信号線00Yを通してパケット200をXBY0に転送する。

【0047】XBY0は、ノード0からのパケット200を、入力ポート0が受け取る。該XBY0は、入力ポート構成情報レジスタ131、133の入力ポート0と対応したビット(1ビット目)を参照し、図10より、該当ビットに“1”が格納されているパーティション0を選択する。次に、パケット200のXBYルーティングフィールド220の値“0011”を確認し、出力ポート2と3を選択する。そこで、パーティション0の出力ポート構成情報レジスタ132の出力ポート2、3に対応するビット(3と4ビット目)を参照し、該当ビットに“1”が格納されているか確認する。図10より、XBY0のパーティション0の出力構成情報レジスタ132は“1010”であり、出力ポート2に対応するビット3は“1”が格納されているが、出力ポート3に対応するビット4は“0”が格納されている。このため、XBY0は、パケット200の転送処理を中止し、エラー情報レジスタ131をオンとする。

【0048】以上の様に、転送パケットのルーティングフィールドを基に、逐次、各ノードの通信ユニットの分割情報レジスタ、また、各クロスバスイッチの構成情報レジスタの情報を参照することにより、他の分割グループにパケットが転送されないことを保証し、相互結合網の分割運転の信頼性が向上する。

【0049】次に、本実施例において、任意のノードからの同一の分割グループ内へのブロードキャストについて説明する。ここでは、ノードから該同一分割グループ内のノード0、8、9にパケットをブロードキャストするものとする。

【0050】ノード1から同一分割グループ内へブロードキャストする際に、ノード1の演算ユニット100は、内部バス110を通して、通信ユニット115にある分割情報レジスタ120の内容“11001010”(図7)を読み出し、当該上位4ビットの値“1100”をXBXルーティングフィールドに、該下位4ビットの値“1010”をXBYルーティングフィールドに格納し、転送するデータをデータフィールドに格納することにより、図2に示されるパケット200をローカルメモリ105に生成する。次に、演算ユニット100は乱数等を発生させた結果、例えば使用する出力ポートを0Xとし、内部バス110を通して通信ユニット115

に、該パケット200の転送指示をする。通信ユニット115は、演算ユニット100からパケット200の転送指示を受けると、内部バス110を通して、ローカルメモリ105から転送すべきパケット200を読み出し、該パケット200のXBXルーティングフィールドの値“1100”から分割情報レジスタ120の上位4ビット中の、XBX出力ポート0、1と対応したビット(0と1ビット目)を参照し、それらのビットに各々“1”が格納されており、且つ、該パケット200のXBYルーティングフィールドの値“1010”から分割情報レジスタ120の下位4ビット中の、XBY出力ポート0、2と対応したビット(1と3ビット目)を参照し、それらのビットに各々1が格納されていることを確認し、指定された出力ポート0Xから出力信号線10Xを通してXBX0に、該読み出したパケット200を転送する。

【0051】なお、ノード1の通信ユニット115は、該転送パケット200のルーティングフィールドの値から分割情報レジスタ120の出力ポートと対応したビットを参照した際に、何れか一つのビットにでも“1”が格納されていなかったならば、パケットの転送処理を中止し、内部バス110を介して演算ユニット100に割り込みを発生させる。または、エラー情報レジスタ121をオンとする。

【0052】XBX0は、ノード1からのパケット200を、入力ポート1が受け取る。該XBX0は、入力ポート構成情報レジスタ131、133の入力ポート1と対応したビット(2ビット目)を参照し、本例では、図10より、該当ビットに1が格納されているパーティション0を選択する。次に、パケット200のXBXルーティングフィールドの値“1100”を参照し、出力ポート0、1を選択する。次に、パーティション0の出力ポート構成情報レジスタ132の出力ポート0、1に対応するビット(1と2ビット目)を参照し、本例では、図10より、該当ビットに“1”が格納されているのを確認する。そこで、XBX0は、出力ポート0、1の出力信号線00X、10Xを通して、該信号線00X、10Xと接続されているノード0、ノード1にパケット200を転送する。

【0053】パケット200を受け取ったノード0、ノード1の各通信ユニットは、該パケットがブロードキャストであることを確認した後、それぞれの出力ポートから出力信号線00Y、10Yを通してパケット200をXB0、XB1に転送する。

【0054】XB0、XB1は、各々ノード0、1からのパケット200を、入力ポート0で受け取る。該XB0、XB1は、各々入力ポート構成情報レジスタ131、133の入力ポート0と対応したビット(1ビット目)を参照し、本例では、該当ビットに“1”が格納されているパーティション0を選択する。次に、XB

BY0、XB1は、各々パケット200のXBYルーティングフィールドの値“1010”を確認し、出力ポート0、2を選択する。次に、XB0、XB1は、それぞれのパーティション0の出力ポート構成情報レジスタ132の出力ポート0、2に対応するビット(1と3ビット目)を参照し、そのビットに1が格納されているのを確認した後、各々出力信号線00Y、80Y、10Y、90Yと接続されているノード0、ノード8、ノード1、ノード9にパケット200を転送する。

【0055】なお、各クロスバスイッチXBX0、XB0、XB1の構成情報レジスタのビットを参照した際に、何れか一つの該当ビットにでも“1”が格納されていなければ、パケットの転送処理は中止し、エラー情報レジスタ131をオンとする。

【0056】以上、本発明の一実施例として、各XB、XBの入出力ポートをパーティション0、パーティション1に2分割する場合について説明したが、本発明は一般的に、XBの入出力ポートをM分割、XBの入出力ポートをN分割する任意の組み合せに適当可能である(ただし、M、Nは1から最大ポート数)。この場合、XB、XBには、MあるいはN組の各入出力ポート構成情報レジスタを設ければよい。

【0057】

【発明の効果】本発明によれば、並列計算機の分割運転時、ノードが分割情報レジスタを基に転送パケットの各ルーティングフィールドの決定や確認を行い、また、クロスバスイッチが構成情報レジスタのビットを基に各ルーティングフィールドの確認を行うことにより、他の分割グループにパケットが転送されないことを保証し、高信頼性の分割運転が可能になる。該相互結合網の分割運転を行っている。

【図面の簡単な説明】

【図1】本発明の一実施例の並列計算機を示す構成図である。

【図2】通信パケットの構成例を示す図である。

【図3】転送先ノード番号とルーティングフィールドの対応を示す図である。

【図4】分割グループとパーティションの対応を示す図である。

【図5】ノードの構成例を示す図である。

【図6】ノード内の分割情報レジスタの構成例を示す図である。

【図7】実施例における分割情報レジスタのビットパターンを示す図である。

【図8】クロスバスイッチの構成例を示す図である。

【図9】クロスバスイッチ内の構成情報レジスタの構成例を示す図である。

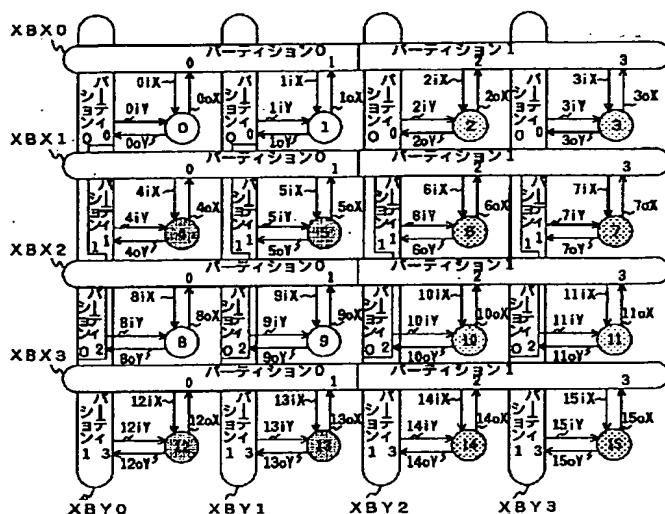
【図10】実施例における構成情報レジスタのビットパターンを示す図である。

【符号の説明】

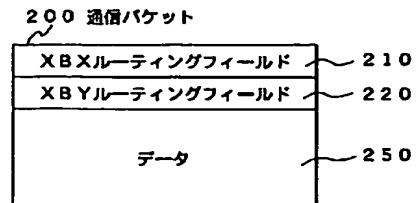
1~15 ノード
XBX0~XBX3、XBY0~XBY3 クロスバス
イッチ
120 分割情報レジスタ

130 構成情報レジスタ
200 通信パケット
210, 220 ルーティングフィールド

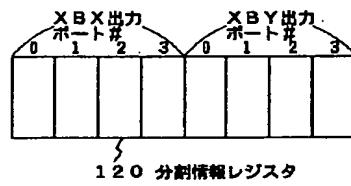
【図1】



【図2】



【図6】



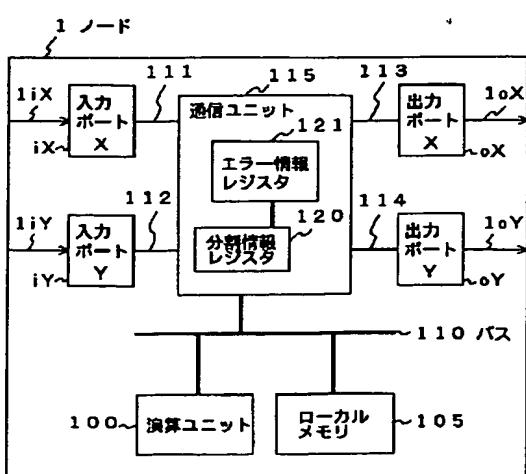
【図4】

分割グループ	ノード番号	パーティション(X, Y)
第1分割グループ	0, 1, 8, 9	(0, 0)
第2分割グループ	2, 3, 10, 11	(1, 0)
第3分割グループ	4, 5, 12, 13	(0, 1)
第4分割グループ	6, 7, 14, 15	(1, 1)

【図3】

ノード番号	2進数	ルーティングフィールド		
		XBX出力ポート# 0 1 2 3	XBY出力ポート# 0 1 2 3	
0	0000	1000	1000	
1	0001	0100	1000	
2	0010	0010	1000	
3	0011	0001	1000	
4	0100	1000	0100	
5	0101	0100	0100	
6	0110	0010	0100	
7	0111	0001	0100	
8	1000	1000	0010	
9	1001	0100	0010	
10	1010	0010	0010	
11	1011	0011	0010	
12	1100	1000	0001	
13	1101	0100	0001	
14	1110	0010	0001	
15	1111	0001	0001	

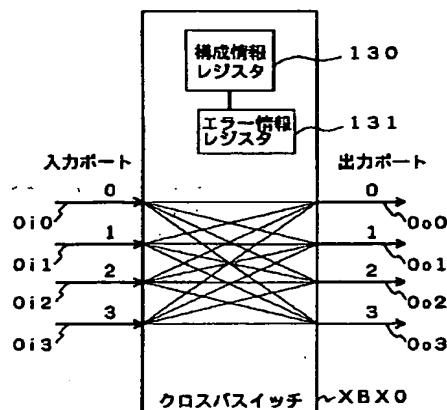
【図5】



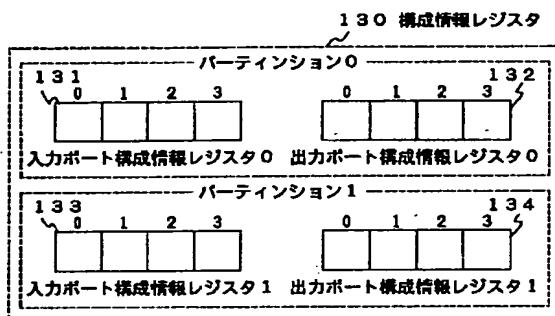
【図 7】

120 分割情報レジスタ							
第1分割グループ	1	1	0	0	1	0	1
第2分割グループ	0	0	1	1	1	0	1
第3分割グループ	1	1	0	0	0	1	0
第4分割グループ	0	0	1	1	0	1	0

【図 8】



【図 9】



【図 10】

パーティション	ポート#				ポート#				
	0	1	2	3	0	1	2	3	
XBX0	1	1	0	0	XBX0	1	1	0	0
XBX1	1	1	0	0	XBX1	1	1	0	0
XBX2	1	1	0	0	XBX2	1	1	0	0
XBX3	1	1	0	0	XBX3	1	1	0	0
XBY0	1	0	1	0	XBY0	1	0	1	0
XBY1	1	0	1	0	XBY1	1	0	1	0
XBY2	1	0	1	0	XBY2	1	0	1	0
XBY3	1	0	1	0	XBY3	1	0	1	0

パーティション 0 の入力ポート構成情報レジスタ (0) 131 と出力ポート構成情報レジスタ (0) 132 の構成:

パーティション 0	ポート# 0	1	2	3	パーティション 0	ポート# 0	1	2	3
131	0	1	2	3	132	0	1	2	3

パーティション 1 の入力ポート構成情報レジスタ (1) 133 と出力ポート構成情報レジスタ (1) 134 の構成:

パーティション 1	ポート# 0	1	2	3	パーティション 1	ポート# 0	1	2	3
133	0	1	2	3	134	0	1	2	3

フロントページの続き

(72) 発明者 舟塙 伸夫

神奈川県秦野市堀山下1番地 株式会社日

立製作所汎用コンピュータ事業部内